

# BEST AVAILABLE COPY

(12) NACH DEM VERTRÄG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES  
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum  
Internationales Büro



(43) Internationales Veröffentlichungsdatum  
3. Januar 2003 (03.01.2003)

PCT

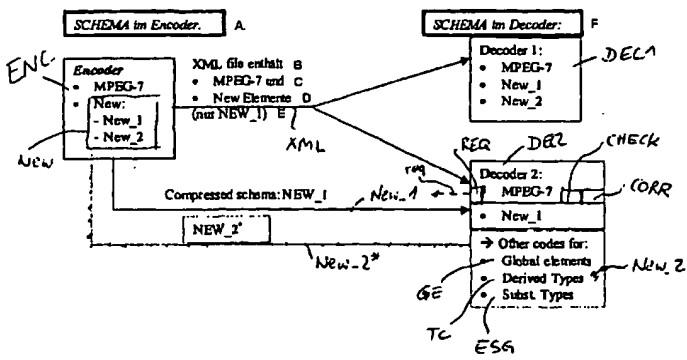
(10) Internationale Veröffentlichungsnummer  
**WO 03/001811 A1**

- |  |  |   |
|--|--|---|
| (51) Internationale Patentklassifikation <sup>7</sup> :  | H04N 7/24                                  | (72) Erfinder; und  |
| (21) Internationales Aktenzeichen:   | PCT/DE02/02309                             | (75) Erfinder/Anmelder (nur für US): HUTTER, Andreas [DE/DE]; Kesselbergstr. 14, 81539 München (DE).                                  |
| (22) Internationales Anmelddatum:  | 25. Juni 2002 (25.06.2002)                 | HEUER, Jörg [DE/DE]; Fischbachauerstr. 8, 81539 München (DE). NIEDERMEIER, Ulrich [DE/DE]; Emanuelstr. 18, 80796 München (DE).        |
| (25) Einreichungssprache:  | Deutsch                                    | (74) Gemeinsamer Vertreter: SIEMENS AKTIENGESELLSCHAFT; Postfach 22 16 34, 80506 München (DE).  |
| (26) Veröffentlichungssprache:   | Deutsch                                    | (81) Bestimmungsstaaten (national): CN, JP, US.   |
| (30) Angaben zur Priorität:  | 101 30 525.7 25. Juni 2001 (25.06.2001) DE | (84) Bestimmungsstaaten (regional): europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR). |
| (71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): SIEMENS AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacherplatz 2, 80333 München (DE). |  |   |

[Fortsetzung auf der nächsten Seite]

(54) Title: SYSTEM FOR THE IMPROVED ENCODING/DECODING OF STRUCTURED, PARTICULARLY XML-BASED, DOCUMENTS AND METHODS AND DEVICES FOR THE IMPROVED ENCODING/DECODING OF BINARY REPRESENTATIONS OF SUCH DOCUMENTS

(54) Bezeichnung: SYSTEM ZUR VERBESSERTEN ENCODIERUNG/DECODIERUNG VON STRUKTURIERTEN, INSbesondere XML-BASIERTEN, DOKUMENTEN SOWIE VERFAHREN UND VORRICHTUNGEN ZUR VERBESSERTEN ENCODIERUNG/DECODIERUNG VON BINÄREN REPRÄSENTATIONEN VON SOLCHEN DOKUMENTEN



- A... SCHEMA IN THE ENCODER:  
B... XML FILE CONTAINS  
C... MPEG-7 AND  
D... NEW ELEMENTS  
E... (ONLY NEW\_1)  
F... SCHEMA IN THE DECODER:

**WO 03/001811 A1**

(57) Abstract: The invention is essentially based on, in the schema definition, which is known to the encoder and to the decoder or transmitted from the encoder to the decoder, specifying information for calculating code concerning the elements or data types contained in the name subspace and enabling this information to be clearly assigned in this manner to each code in the name subspace. Only a portion of the entire name space including the number of the elements or data types lacking for the code calculation have to be transmitted for the configuration of a code which generally means a significantly smaller volume of data than the volume of data that would be required for the entire name space. In addition, the code tables are smaller than in entirely known name spaces which requires less storage space in the codec and enables a faster encoding and decoding.

[Fortsetzung auf der nächsten Seite]

**Erklärungen gemäß Regel 4.17:**

- hinsichtlich der Berechtigung des Anmelders, ein Patent zu beantragen und zu erhalten (Regel 4.17 Ziffer ii) für die folgenden Bestimmungsstaaten CN, JP, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR)
- vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eintreffen

**Veröffentlicht:**

- mit internationalem Recherchenbericht

Zur Erklärung der Zweibuchstaben-Codes und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

---

**(57) Zusammenfassung:** Die Erfindung besteht im Wesentlichen darin, dass in der Schemadefinition, die dem Encoder und dem Decoder bekannt ist oder vom Encoder an den Decoder übermittelt wird, Informationen für die Codeberechnung über die nicht im Namensraum enthaltenen Elemente bzw. Datentypen mitangegeben werden und auf diese Weise jeder Code im Namensraum eindeutig zugewiesen werden kann. Dadurch muss für die Konfiguration eines Codecs nur ein Teil des gesamten Namensraums inklusive der Anzahl der für die Codeberechnung fehlenden Elemente bzw. Datentypen übertragen werden, was im Allgemeinen ein wesentlich geringeres Datenvolumen bedeutet, als das Datenvolumen, das für den gesamten Namensraum benötigt würde. Zudem sind die Codetabellen kleiner als bei vollständig bekannten Namensräumen, was weniger Speicherplatz im Codec erfordert und eine schnellere En- und Decodierung ermöglicht.

**Beschreibung**

System zur verbesserten Encodierung/Decodierung von strukturierten, insbesondere XML-basierten, Dokumenten sowie  
5 Verfahren und Vorrichtung zur verbesserten Encodierung/Decodierung von binären Repräsentationen von solchen Dokumenten

Die Erfindung betrifft Verfahren, Vorrichtungen oder Systeme,  
10 bei denen eine binäre Repräsentation eines strukturierten, insbesondere XML-basierten Dokuments mit Hilfe eines Schemas codiert und/oder decodiert wird.

Derartige Verfahren, Vorrichtungen oder Systeme sind  
15 beispielsweise aus Schriften zum MPEG-7-Standard, insbesondere aus dem "Text of ISO/IEC FCD 15938-1 Information Technology - Multimedia Content Description Interface - Part1, Systems" der Systems Sub-Group, von Claude Seyrat (Expway), Michael Wollborn (Bosch), Ali Tabatabai (Sony,  
20 Olivier Avaro (France Telecom R&D) bzw. ISO/IEC JTC1/SC29/WG11, MPEG 01/N4001, March 2001, Singapore", bekannt.

Diese Verfahren zur binären Repräsentation von MPEG-7 und  
25 anderen XML-basierten Beschreibungen oder Dokumenten weisen Defizite hinsichtlich der Kompatibilität auf, sofern einem Decoder nur ein Teil des gesamten Namensraums und/oder eines Schemas, der zur Codierung genutzt wurde, bekannt ist.

30 Die der Erfindung zugrundeliegende Aufgabe besteht nun darin, die oben genannten Nachteile zu vermeiden.

Diese Aufgabe wird hinsichtlich eines besseren Verfahrens zur Decodierung durch die Merkmale des Patentanspruchs 1,  
35 hinsichtlich besseren Verfahrens zur Encodierung durch die Merkmale des Patentanspruchs 8,

- hinsichtlich eines Systems zur verbesserten  
Encodierung/Decodierung durch die Merkmale des  
Patentanspruchs 15, hinsichtlich einer Vorrichtung zur  
verbesserten Decodierung durch die Merkmale des  
5 Patentanspruchs 17 und hinsichtlich einer Vorrichtung zur  
verbesserten Encodierung durch die Merkmale des  
Patentanspruchs 18 gelöst. Die weiteren Ansprüche beziehen  
sich auf vorteilhafte Ausgestaltungen der Verfahren und des  
Systems.

10

- Die Erfindung besteht im Wesentlichen darin, dass in der  
Schemadefinition, die dem Encoder und dem Decoder bekannt ist  
oder vom Encoder an den Decoder übermittelt wird,  
Informationen für die Codeberechnung über die nicht im  
15 Namensumerraum enthaltenen Elemente bzw. Datentypen  
mitangegeben werden und auf diese Weise jeder Code in einem,  
im Folgenden als Namensumerraum bezeichneten, Teil des  
gesamten Namensraums eindeutig zugewiesen werden kann.  
Dadurch muss für die Konfiguration eines Codecs nur ein Teil  
20 des gesamten Namensraums inklusive der Anzahl der für die  
Codeberechnung fehlenden Elemente bzw. Datentypen übertragen  
werden, was im Allgemeinen ein wesentlich geringeres  
Datenvolumen bedeutet, als das Datenvolumen, das für den  
gesamten Namensraum benötigt würde. Zudem sind die  
25 Codetabellen kleiner als bei vollständig bekannten  
Namensräumen, was weniger Speicherplatz im Codec erfordert  
und eine schnellere En- und Decodierung ermöglicht.
- Die Erfindung wird nachfolgend anhand von Zeichnungen zu  
30 Ausführungsbeispielen der Erfindung näher erläutert. Dabei  
zeigt

Figur 1 eine Prinzipdarstellung eines erfindungsgemäßen  
Systems mit Encoder und Decoder,

35

Figur 2 einen Bitstrom eines Korrekturcodes von Figur 1,

Figur 3 eine Darstellung zur Erläuterung der Korrektur bei  
Globalelementen,

Figur 4 eine Darstellung zur Erläuterung der Korrektur bei  
5 Elementen einer Ersetzungsgruppe und

Figur 5 eine Darstellung zur Erläuterung der Korrektur bei  
Typidentifikationscodes.

10

In Figur 1 ist beispielhaft ein System mit einem Encoder ENC und zwei Decodern DEC1 und DEC2 dargestellt, wobei dem Decoder ein Schema MPEG-7 sowie ein Schema New mit den Namensunterräumen NEW\_1 und New\_2 zur Verfügung steht. Der 15 Encoder sendet hier ein XML-File XML, das MPEG-7-Elemente und Elemente des Schemas New enthält, wobei letztere Elemente aber nur aus dem Namensraum NEW\_1 stammen. Der Decoder DEC1, dem beide Namensräume MPEG-7 und New bzw. die Namensunterräume New\_1 und New\_2 bekannt sind, kann dieses 20 Dokument natürlich in entsprechender Weise decodieren. Dem Decoder DEC2 ist hingegen nur der Namensraum MPEG-7 bekannt. Daher wird diesem, zum Beispiel auf ein Anforderungssignal req hin oder zum Beispiel durch regelmäßige Übermittlung, der Namensraum New\_1 als Schema übermittelt. Allerdings kann 25 der Decoder DEC2 nun das Dokument immer noch nicht decodieren, da ihm die Codezuweisung von globalen Elementen GE, Typcodes TC und Substitution Codes ESG aus dem Unterraum New\_2 fehlen und er somit die Codezuweisungen nicht berechnen kann. Um dieses Problem zu beseitigen, wird nicht der gesamte 30 Namensraum New\_2 dem Decoder DEC2 zur Verfügung gestellt, sondern vorteilhafter Weise nur ein Korrekturcode New\_2\* an Stelle der im Namensraum und/oder vereinfachten Schema New\_1 nicht enthaltenen Elemente und/oder Typen des vollständigen Namensraums übertragen. Aus 35 dem übermittelten vereinfachten Schema kann nun der Decoder DEC2 mit Hilfe des Korrekturcodes das XML-basierte Dokument

XML gemäß dem vollständigen Namensraum und/oder Schema decodieren.

In Figur 2 ist ein Bitstrom für den Korrekturcode New\_2\*  
5 gezeigt, der mindestens ein Triple aus einem Tabellentyp TT  
der zu korrigierenden Elemente GE, ESG oder Typen TC, einer  
Tabellenidentifizierung TID zur Festlegung der zu  
korrigierenden Tabelle und einer Korrekturinformation MAP,  
GAP oder MX aufweist. Aus der durch den Tabellentyp und die  
10 Tabellenidentifizierung festgelegten Codetabelle des  
vereinfachten Schemas New\_1 wird nun mit Hilfe der  
Korrekturinformation New\_2\* eine Codetabelle des  
vollständigen Schemas New erzeugt, die nur Einträge für die  
Elemente und Typen aus dem Namensumerraum und/oder  
15 vereinfachten Schema New\_1 enthält.

Der Tabellentyp TT spezifiziert, ob ein globales Element GE,  
Typcode TC oder Elemente einer Substitutionsgruppe ESG  
ergänzt wird. Die Tabellenidentifikation TID identifiziert  
20 die zu manipulierende Tabelle, wobei dies beim Typcode mit  
dem Typcode des Basistypen bezogen auf den Urtyp und bei  
einer Substitutionsgruppe mit dem Schemaverzweigungscode SBC  
des globalen Kopfelementes erfolgt. Bei globalen Elementen  
wird das Feld TID hingegen nicht codiert.  
25

In Figur 3 wird beispielhaft die Codierung der globalen  
Elemente GE(New\_1) im Decoder DEC2 und die Codierung der  
globalen Elemente GE(New) im Encoder ENC, also im  
vollständigen Namensraum, miteinander verglichen und die  
30 Unterschiede in den Tabellencodes festgestellt. Hieraus wird  
deutlich, dass die Codes der Elemente 1 bis 3 im Decoder DEC2  
mit 00, 01 und 10 im Encoder ENC hingegen mit 0001, 0010 und  
0110, dass also beispielsweise Lücken 0000 zu Beginn, Lücken  
0011...0101 in der Mitte und Lücken 0111 am Ende der  
35 Codetabelle auftreten können.

- In Figur 4 ist entsprechenderweise die Codierung von Elementen ESG(New\_1) von Substitutionsgruppen im Decoder DEC2, also eine Decodierung im Namensunerraum New\_1, und eine Codierung der Elemente ESG(New) dieser
- 5 Substitutionsgruppe im vollständigen Namensraum New sowie die zugehörigen Tabellencodes dargestellt. Auch hier wird deutlich, das die Codes 00, 01 und 10 der Elemente 1 bis 3 des Decoders DEC2 Codes im Encoder entsprechen, die an beliebiger Stelle der entsprechenden Codetabelle liegen
- 10 können.

In Figur 5 wird dieser Vergleich zwischen den Typidentifikationscodes TC(New\_1) im Decoder DEC2 und den Typidentifikationscodes TC(New) für denselben Basistyp im Encoder ENC, also im vollständigen Namensraum dargestellt. Die abgeleiteten Typen im Namensunerraum, denen Typidentifikationscodes TC(New\_1) zugeordnet sind, können hier beispielsweise in einer baumartige Struktur angeordnet werden, wobei mit einem Wurzelknoten 0 der Basistyp symbolisiert wird, der Kindknoten 1 und 4 aufweist, die vom Basistyp direkt abgeleitete Typen repräsentieren, wobei der Kindknoten 1 wiederum Kindknoten 2 und 3 und der Kindknoten 4 einen Kindknoten 5 aufweist, die von abgeleiteten Typen wiederum abgeleitete Typen repräsentieren. Die fünf Typen 0 bis 5 im Decoder DEC2 werden im Encoder in einer baumartigen Struktur mit neun Knoten 0 bis 8 repräsentiert. Die baumartige Struktur mit dem vollständigen Namensraum weist einen Wurzelknoten 0 und drei Kindknoten 1, 5 und 6 auf, wobei der Kindknoten 1 wiederum drei Kindknoten 2, 3 und 4 und der Kindknoten 6 zwei Kindknoten 7 und 8 beisitzt. Dementsprechend sind die Datentypen Nummer 2, 5 und 8 dem Decoder DEC2 aus dem Namensunerraum nicht bekannt. Als eindeutiges Codezuordnungsprinzip kann beispielsweise die sogenannte "Depth first"-Methode nach lexikographischer

25 Ordnung nach dem zitierten MPEG-7 Standard verwendet werden, das also zuerst die möglichen Knoten in der Tiefe bzw. in Richtung Blätter einer baumartigen Struktur adressiert, um

- die baumartige Struktur der abgeleiteten Typen in eine Sequenz an Typen abzubilden und den Typen in dieser Reihenfolge der Sequenz Typcodes zuzuweisen. Dementsprechend können die Codes auch Tabellarisch entsprechend der globalen
- 5 Elemente und entsprechend einer Substitutionsgruppe dargestellt werden.

Das letzte Feld MAP, MIX, GAP eines Triples im Bitstrom von Figur 2 enthält schließlich Informationen über Elemente bzw.

10 Typen in einer jeweiligen Tabelle bzw. einem jeweiligen Baum, die nicht übertragen worden sind.

Diese Informationen können entsprechend dreier Ausgestaltungen des erfindungsgemäßen Verfahrens codiert

15 werden, die im Folgenden näher beschrieben werden:

Im Falle der Methode MAP wird zu jedem Code, der aufgrund der ausschließlichen Kenntnis des Namensraums New\_1 vom Decoder DEC2 nicht entsprechend dem Encoder ENC berechnet

20 wird, ein Korrekturcode übertragen, der zwei Teile enthält, nämlich eine Länge der Codewörter des vollständigen Namensraums und/oder Schemas und eine Liste mit Codeersetzungen. Der Decoder DEC2 ersetzt dabei die in der adressierten Codetabelle bzw. im Codebaum enthaltenen Codes

25 mit den übertragenen Codes. Dieses Verfahren eignet sich besonderes dann, wenn der Namensraum New\_1 nur wenige Elementdeklarationen bzw. Typdefinitionen enthält.

Im Falle der Methode GAP wird ein Korrekturcode übertragen, der nur aus einer Liste mit Lückenzahlen für die Codetabelle des vollständigen Namensraums und/oder Schemas enthält. Der Decoder DEC2 liest n+1 Zahlen ein, wobei n die Anzahl der Einträge in der adressierten Codetabelle bzw. dem Codebaum des Namensraums New\_1 ist. Entsprechend der eingelesenen Zahl werden an der jeweiligen Stelle in der Codetabelle bzw. dem Codebaum Einträge generiert. Anschließend werden die Codes aller Einträge neu berechnet. Dieses Verfahren eignet

sich besonders dann, wenn der Namensumtraum New\_1 nur wenige Elementdeklarationen bzw. Datentypdefinitionen enthält.

- 5    Im dritten Fall MIX enthält die Korrekturinformation drei Teile, nämlich eine Länge der Codewörter des vollständigen Namensraums und/oder Schemas, eine Anzahl der Lücken in der Codetabelle in Bezug auf den vollständigen Namensraum und/oder Schema und eine Liste mit Codeersetzungen, wobei
- 10   jeweils eine Codeersetzung für ein Element oder einen Typ direkt nach einer Lücke in der Codetabelle durchgeführt wird und anschließend die Codezuweisung aller restlichen Einträge in der Codetabelle durchgeführt wird. Der Decoder DEC2 liest den neuen Code des Eintrags ein und modifiziert die Codes der
- 15   folgenden Einträge entsprechend des eingangs genannten bekannten Verfahrens. Dieses Verfahren eignet sich besonders für viele fehlende Elementdeklarationen bzw. Typdefinitionen, die zusammenhängen.
- 20   Alle drei im Vorangehenden beschriebene Ausgestaltungen können für die Korrekturinformation wahlweise genutzt werden, indem vor der eigentlichen Korrekturinformation beispielsweise mit zwei Bit die im folgenden genutzte Ausgestaltung der Korrekturinformation signalisiert wird.

## Patentansprüche

1. Verfahren zur verbesserten Decodierung von binären Repräsentationen von XML-basierten Dokumenten,
  - 5 bei dem mindestens ein Decoder zur Decodierung den vollständigen Namensraum und/oder Schema (new) benötigt und dem ein Namensunterraum und/oder vereinfachtes Schema (New\_1) mit den Elementen und/oder Typen des Dokuments als vereinfachtes Schema übermittelt wird,
  - 10 bei dem zusätzlich nur ein Korrekturcode (New\_2\*) an Stelle der im Namensunterraum und/oder vereinfachten Schema (New1) nicht enthaltenen Elemente und/oder Typen (New\_2) des vollständigen Namensraums und/oder Schemas dem mindestens einen Decoder übermittelt wird und
  - 15 bei dem der Decoder aus dem übermittelten vereinfachten Schema mit Hilfe des Korrekturcodes das XML-basierte Dokument gemäß dem vollständigen Namensraum und/oder Schema decodiert.
- 20 2. Verfahren nach Anspruch 1,
  - 25 bei dem der mindestens eine Decoder, falls dieser vollständige Namensraum und/oder Schema in dem mindestens einen Decoder nicht vorhanden ist, ein Anfragesignal (req) zur Übermittlung des Namensunterraums und/oder vereinfachten Schemas und des Korrekturcodes bildet.
- 30 3. Verfahren nach Anspruch 1 oder 2,
  - 35 bei dem der Korrekturcode aus einem Bitstrom mit mindestens einem Tripel aus einem Tabellentyp (TT) der zu korrigierenden Elemente (GE, ESG) oder Typen (TC), einer Tabellenidentifizierung (TID) zur Festlegung der zu korrigierenden Tabelle, wobei bei globalen Elementen (GE) keine Codierung erfolgt, und einer Korrekturinformation (Code) besteht.
- 40 4. Verfahren nach einem der Ansprüche 1 bis 3, bei dem aus einer durch den Tabellentyp und die Tabellenidentifizierung festgelegten Codetabelle des

Namensunerraums und/oder vereinfachten Schemas mit Hilfe der Korrekturinformation eine Codetabelle des vollständigen Namensraums und/oder Schemas erzeugt wird.

- 5    5. Verfahren nach einem der Ansprüche 1 bis 4, bei dem die Korrekturinformation (MAP) eine Länge der Codewörter des vollständigen Namensraums und/oder Schemas und eine Liste mit Codeersetzungen enthält.
- 10    6. Verfahren nach einem der Ansprüche 1 bis 4, bei dem die Korrekturinformation (GAP) eine Liste mit Lückenanzahlen für die Codetabelle des vollständigen Namensraums und/oder Schemas enthält.
- 15    7. Verfahren nach einem der Ansprüche 1 bis 4, bei dem die Korrekturinformation (MIX) eine Länge der Codewörter des vollständigen Namensraums und/oder Schemas, eine Anzahl der Lücken in der Codetabelle in Bezug auf den vollständigen Namensraum und/oder Schema und eine Liste mit Codeersetzungen enthält, wobei jeweils eine Codeersetzung für ein Element oder einen Typ direkt nach einer Lücke in der Codetabelle durchgeführt wird und abschließend die Codezuweisung aller restlichen Einträge in Codetabelle durchgeführt wird.
- 25    8. Verfahren zur verbesserten Encodierung von XML-basierten Dokumenten, bei dem ein aus einem XML-basierten Dokument mit mindestens einem vollständigen Namensraum und/oder Schema (new) eine binäre Repräsentation des XML-basierten Dokuments gebildet und übermittelt wird, bei dem ein Namensunerraum und/oder vereinfachtes Schema (New\_1) mit den Elementen und/oder Typen des Dokuments als vereinfachtes Schema und zusätzlich nur ein Korrekturcode (New\_2\*) an Stelle der im Namensunerraum und/oder vereinfachten Schema (New\_1) nicht enthaltenen Elemente und/oder Typen (New\_2) des vollständigen Namensraums und/oder Schemas gebildet und übermittelt wird.

10

9. Verfahren nach Anspruch 8,  
bei dem der Encoder erst auf ein Anfragesignal (req) hin,  
einen Namensumerraum und/oder Schema (New\_1) mit den  
Elementen und/oder Typen des Dokuments als vereinfachtes  
Schema und zusätzlich nur einen Korrekturcode (New\_2\*) an  
Stelle der im Namensumerraum und/oder vereinfachten Schema  
nicht enthaltenen Elementen und/oder Typen (New\_2) des  
vollständigen Namensraums und/oder Schemas bildet und  
übermittelt.
- 10
10. Verfahren nach Anspruch 8 oder 9,  
bei dem der Korrekturcode aus einem Bitstrom mit mindestens  
einem Tripel aus einem Tabellentyp (TT) der zu korrigierenden  
Elemente (GE, ESG) oder Typen (TC),  
15 einer Tabellenidentifizierung (TID) zur Festlegung der zu  
korrigierenden Tabelle, wobei bei globalen Elementen (GE)  
keine Codierung erfolgt, und  
einer Korrekturinformation (Code) besteht.
- 20 11. Verfahren nach einem der Ansprüche 8 bis 10,  
bei dem aus einer durch den Tabellentyp und die  
Tabellenidentifizierung festgelegten Codetabelle des  
Namensumerraums und/oder vereinfachten Schemas mit Hilfe der  
Korrekturinformation eine Codetabelle des vollständigen  
25 Namensraums und/oder Schemas erzeugt wird.
12. Verfahren nach einem der Ansprüche 8 bis 11,  
bei dem die Korrekturinformation (MAP) eine Länge der  
Codewörter des vollständigen Namensraums und/oder Schemas und  
30 eine Liste mit Codeersetzungen enthält.
13. Verfahren nach einem der Ansprüche 8 bis 11,  
bei dem die Korrekturinformation (GAP) eine Liste mit  
Lückenanzahlen für die Codetabelle des vollständigen  
35 Namensraums und/oder Schemas enthält.

14. Verfahren nach einem der Ansprüche 8 bis 11,  
bei dem die Korrekturinformation (MIX) eine Länge der  
Codewörter des vollständigen Namensraums und/oder Schemas,  
eine Anzahl der Lücken in der Codetabelle in Bezug auf den  
5 vollständigen Namensraum und/oder Schema und eine Liste mit  
Codeersetzungen enthält, wobei jeweils eine Codeersetzung für  
ein Element oder einen Typ direkt nach einer Lücke in der  
Codataelle durchgeführt wird und abschließend die  
Codezuweisung aller restlichen Einträge in Codetabelle  
10 durchgeführt wird.
15. System zur verbesserten Encodierung/Decodierung von XML-  
basierten Dokumenten,  
bei dem ein Encoder aus einem XML-basierten Dokument mit  
15 mindestens einem vollständigen Namensraum und/oder Schema  
(new) eine binäre Repräsentation des XML-basierten Dokuments  
bildet an mindestens einen Decoder übermittelt,  
und bei dem ein Encoder einen Namensunterraum und/oder  
vereinfachtes Schema (New\_1) mit den Elementen und/oder Typen  
20 des Dokuments als vereinfachtes Schema und zusätzlich nur  
einen Korrekturcode (New\_2\*) an Stelle der im Namensunterraum  
und/oder vereinfachten Schema (New\_1) nicht enthaltenen  
Elemente und/oder Typen (New\_2) des vollständigen Namensraums  
und/oder Schemas bildet und dem mindestens einen Decoder  
25 übermittelt,  
und bei dem der Decoder aus dem übermittelten vereinfachten  
Schema mit Hilfe des Korrekturcodes das XML-basierte Dokument  
gemäß dem vollständigen Namensraum und/oder Schema decodiert.
- 30 16. Ein System nach Anspruch 15,  
bei dem der mindestens eine Decoder das Vorhandensein dieses  
erforderlichen vollständigen Namensraums und/oder Schemas im  
jeweiligen Decoder überprüft und, falls dieser Vollständige  
Namensraum und/oder Schema in dem mindestens einen Decoder  
35 nicht vorhanden ist, ein Anfragesignal (req) an den Encoder  
bildet,

## 12

- bei dem der Encoder erst auf das Anfragesignal hin, einen Namensumerraum und/oder vereinfachtes Schema (New\_1) mit den Elementen und/oder Typen des Dokuments als vereinfachtes Schema und zusätzlich nur einen Korrekturcode (New\_2\*) an
- 5 Stelle der im Namensumerraum und/oder vereinfachten Schema nicht enthaltenen Elemente und/oder Typen (New\_2) des vollständigen Namensraums und/oder Schemas bildet und dem mindestens einen Decoder übermittelt.
- 10 17. Vorrichtung zur verbesserten Decodierung von binären Repräsentationen von XML-basierten Dokumenten, bei der eine Einheit derart vorhanden ist, dass ein Namensumerraum und/oder vereinfachtes Schema (New\_1) mit den Elementen und/oder Typen des Dokuments als vereinfachtes Schema empfangen wird und zusätzlich nur ein Korrekturcode (New\_2\*) für die im Namensumerraum und/oder vereinfachten Schema nicht enthaltenen Elemente und/oder Typen (New\_2) des vollständigen Namensraums und/oder Schemas empfangen werden und
- 15 20 bei dem eine weitere Einheit derart vorhanden ist, dass aus dem übermittelten vereinfachten Schema mit Hilfe des Korrekturcodes das XML-basierte Dokument gemäß einem vollständigen Namensraum und/oder Schema decodiert wird.
- 25 18. Vorrichtung zur verbesserten Encodierung von binären Repräsentationen von XML-basierten Dokumenten, bei der eine Einheit derart vorhanden ist, dass aus einem XML-basierten Dokument mit mindestens einem vollständigen Namensraum und/oder Schema (new) eine binäre Repräsentation des XML-basierten Dokuments gebildet und gesendet wird, bei der eine weitere Einheit derart vorhanden ist, dass ein Namensumerraum und/oder vereinfachtes Schema (New\_1) mit den Elementen und/oder Typen des Dokuments als vereinfachtes Schema gesendet und zusätzlich nur ein Korrekturcode (New\_2\*) an Stelle der im Namensumerraum und/oder vereinfachten Schema (New\_1) nicht enthaltenen Elemente und/oder Typen

13

(New\_2) des vollständigen Namensraums und/oder Schemas  
gesendet werden.

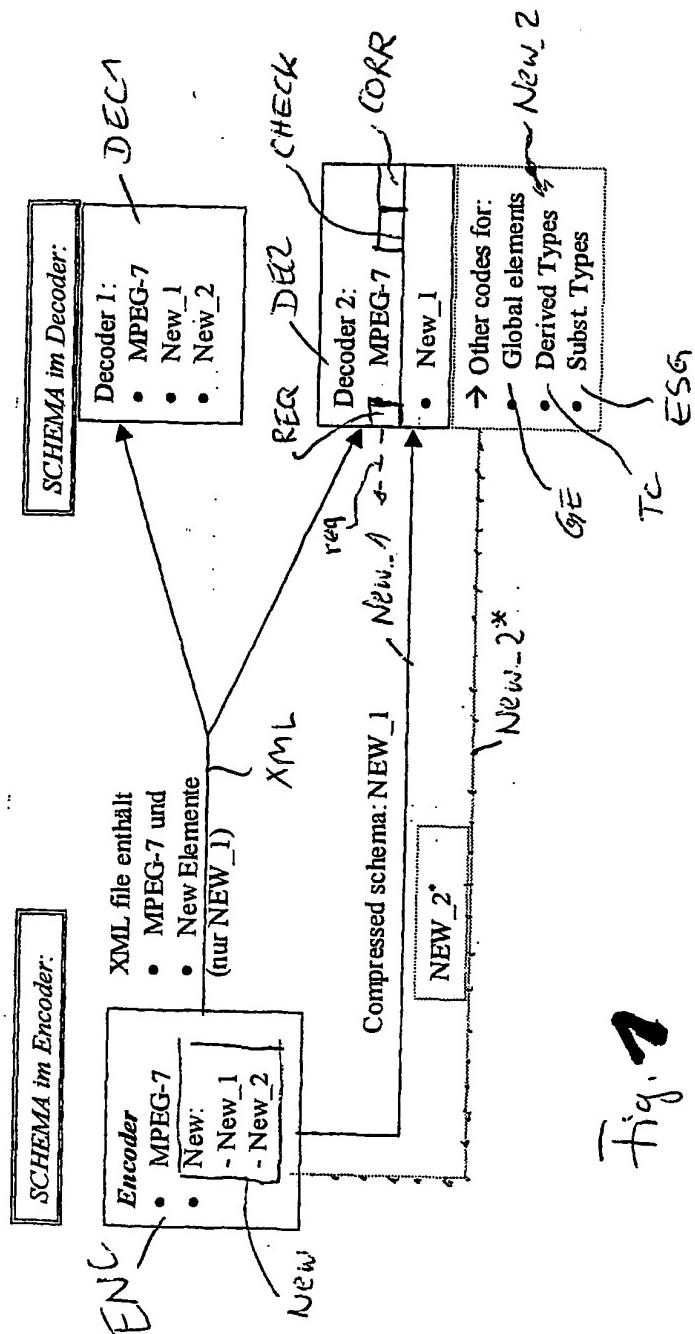
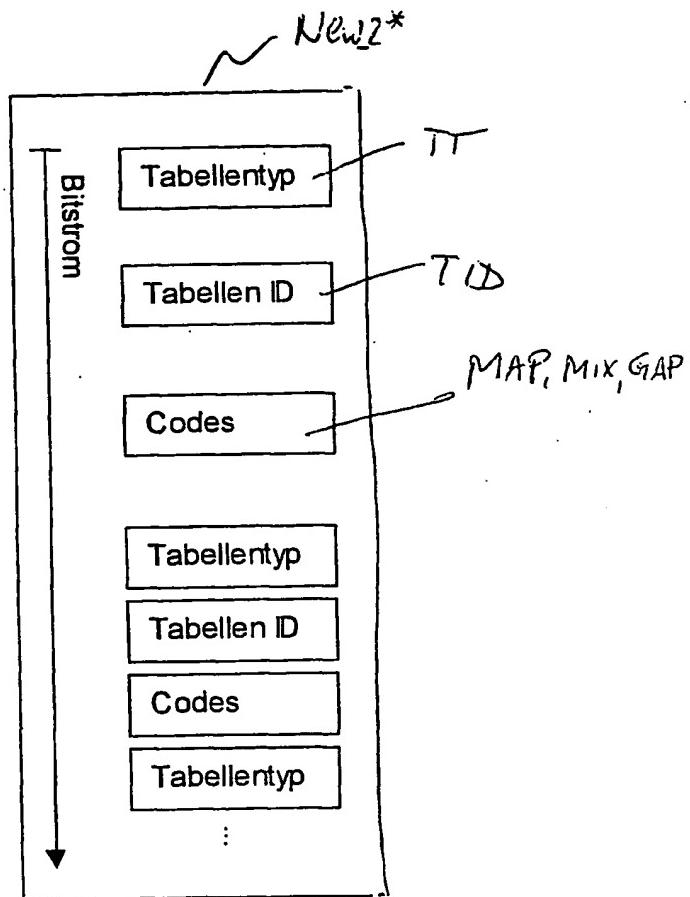


Fig. 1



*Fig. 2*

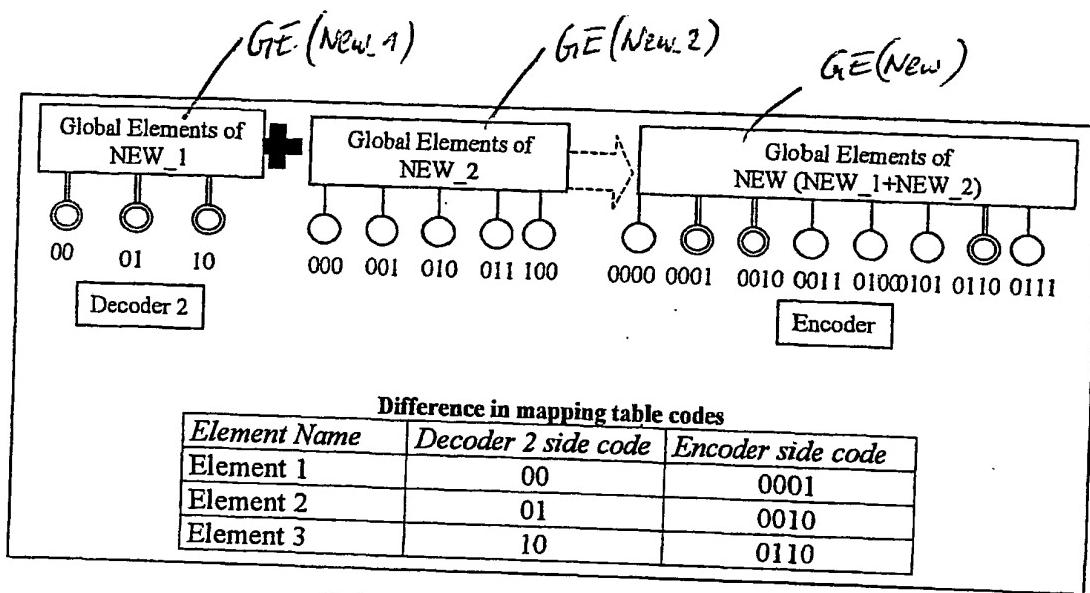


Fig. 3

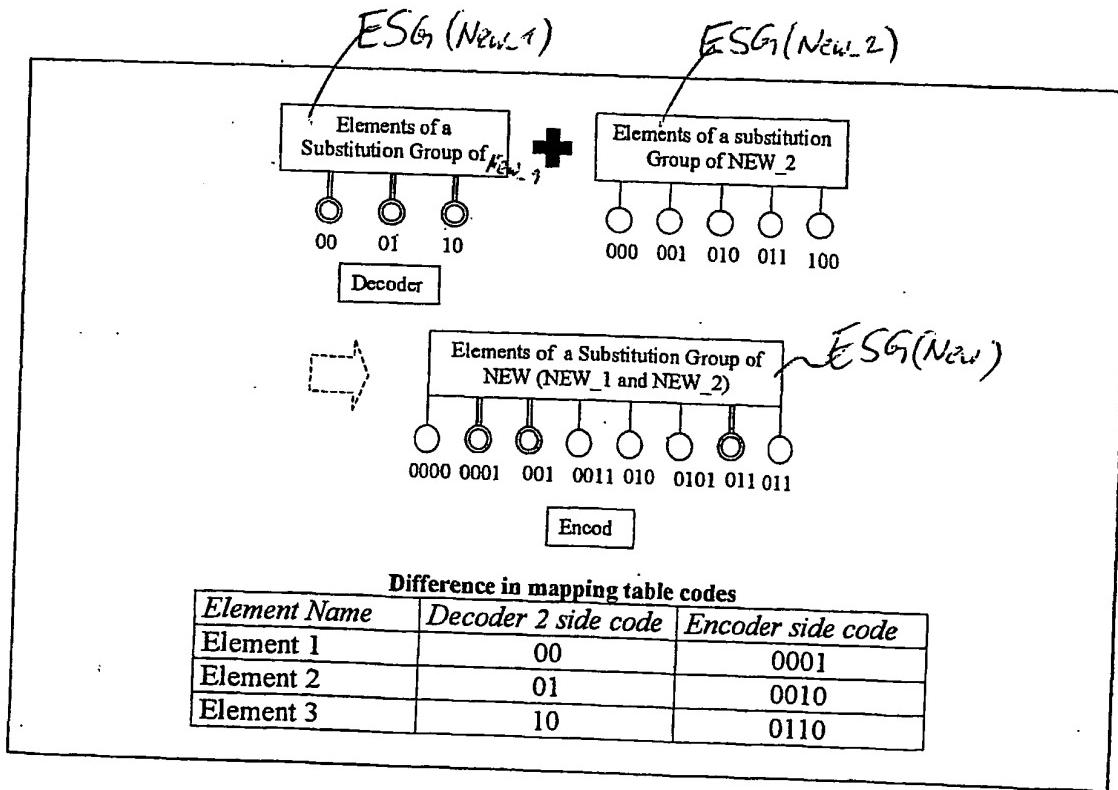
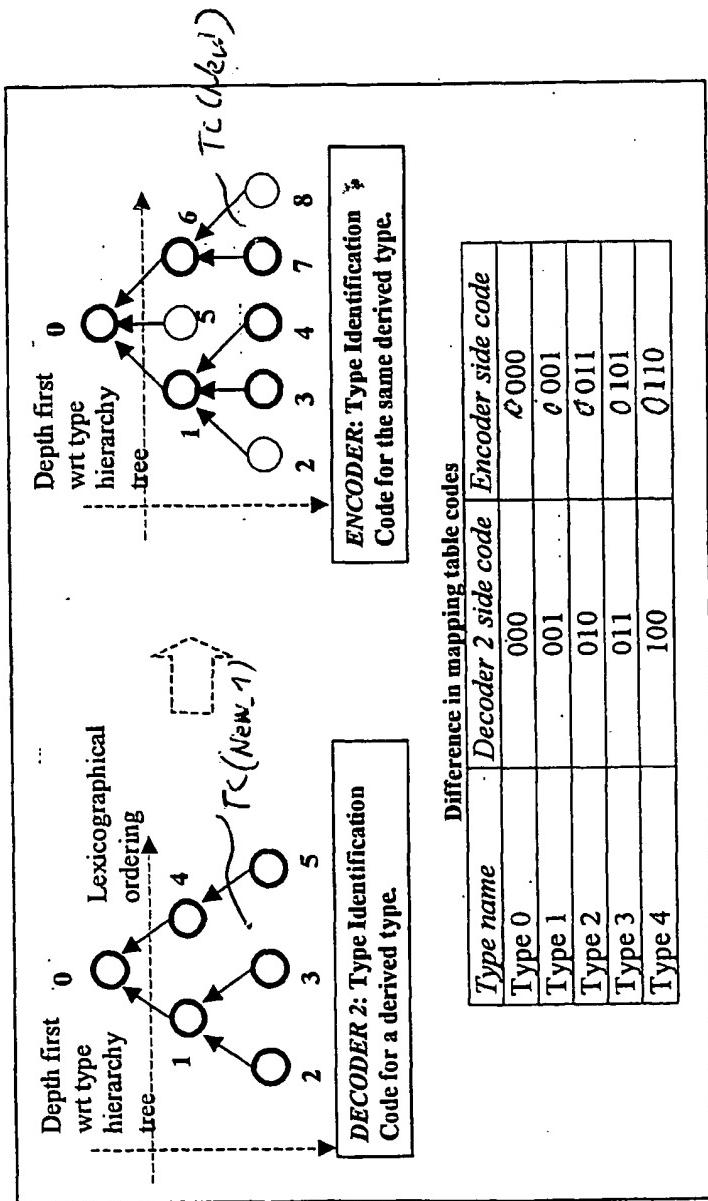


Fig. 4



# INTERNATIONAL SEARCH REPORT

I      National Application No  
 PCT/DE 02/02309

**A. CLASSIFICATION OF SUBJECT MATTER**  
**IPC 7 H04N7/24**

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
**IPC 7 H04N**

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)  
**EPO-Internal, INSPEC**

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 02 063775 A (SEYRAT CLAUDE ;EXPWAY (FR); THIENOT CEDRIC (FR)) 15 August 2002 (2002-08-15) abstract page 2, line 35 -page 3, line 26 page 12, line 35 -page 12, line 38	1,5,8, 12,15, 17,18
A	"TEXT OF ISO/IEC FCD 15938-1 INFORMATION TECHNOLOGY - MULTIMEDIA CONTENT DESCRIPTION INTERFACE - PART 1 SYSTEMS" ISO/IEC JTC1/SC29/WG11 MPEG01/N4001, XX, XX, March 2001 (2001-03), pages 1-2,I-V,6-58, XP001001465 page 55, section 2.8	1,8,15, 17,18

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

° Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

**20 November 2002**

Date of mailing of the international search report

**27/11/2002**

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
 NL - 2280 HV Rijswijk  
 Tel. (+31-70) 340-2040, Tx. 31 651 epo nl  
 Fax (+31-70) 340-3016

Authorized officer

**Hampson, F**

## INTERNATIONAL SEARCH REPORT

Initial Application No
PCT/DE 02/02309

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 34240 A (UNIV MASSACHUSETTS) 18 September 1997 (1997-09-18) abstract page 19, line 8 -page 20, line 28 -----	1,8,15, 17,18

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International Application No

PCT/DE 02/02309

Patent document cited in search report	Publication date		Patent family member(s)		Publication date
WO 02063775	A 15-08-2002	WO	02063775 A2		15-08-2002
WO 9734240	A 18-09-1997	AU WO	2585797 A 9734240 A1		01-10-1997 18-09-1997

## INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/DE 02/02309

A. KLASIFIZIERUNG DES ANMELDUNGSGEGENSTANDES  
IPK 7 H04N7/24

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

## B. RECHERCHIERTE GEBIETE

Recherchierte Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)  
IPK 7 H04N

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, INSPEC

## C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
E	WO 02 063775 A (SEYRAT CLAUDE ;EXPWAY (FR); THIENOT CEDRIC (FR)) 15. August 2002 (2002-08-15) Zusammenfassung Seite 2, Zeile 35 -Seite 3, Zeile 26 Seite 12, Zeile 35 -Seite 12, Zeile 38 —	1,5,8, 12,15, 17,18
A	"TEXT OF ISO/IEC FCD 15938-1 INFORMATION TECHNOLOGY - MULTIMEDIA CONTENT DESCRIPTION INTERFACE - PART 1 SYSTEMS" ISO/IEC JTC1/SC29/WG11 MPEG01/N4001, XX, XX, März 2001 (2001-03), Seiten 1-2,I-V,6-58, XP001001465 page 55, section 2.8 —	1,8,15, 17,18 —/—



Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen



Siehe Anhang Patentfamilie

- \* Besondere Kategorien von angegebenen Veröffentlichungen :
- \*A\* Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
- \*E\* Älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
- \*L\* Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
- \*O\* Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Aussetzung oder andere Maßnahmen bezieht
- \*P\* Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
- \*T\* Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
- \*X\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erforderlicher Tätigkeit beruhend betrachtet werden
- \*Y\* Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erforderlicher Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
- \*&\* Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absendedatum des Internationalen Recherchenberichts
20. November 2002	27/11/2002
Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5018 Patentlaan 2 NL - 2280 HV Rijswijk Tel (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016	Bevollmächtigter Bediensteter  Hampson, F

## INTERNATIONALER RECHERCHENBERICHT

I  
tales Aktenzeichen  
PCT/DE 02/02309

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	WO 97 34240 A (UNIV MASSACHUSETTS) 18. September 1997 (1997-09-18) Zusammenfassung Seite 19, Zeile 8 -Seite 20, Zeile 28	1, 8, 15, 17, 18

**INTERNATIONALER RECHERCHENBERICHT**

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

PCT/DE 02/02309

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
WO 02063775 A	15-08-2002	WO 02063775 A2	15-08-2002
WO 9734240 A	18-09-1997	AU 2585797 A WO 9734240 A1	01-10-1997 18-09-1997

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- BLACK BORDERS**
- IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- FADED TEXT OR DRAWING**
- BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- SKEWED/SLANTED IMAGES**
- COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- GRAY SCALE DOCUMENTS**
- LINES OR MARKS ON ORIGINAL DOCUMENT**
- REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**